```
% Example of Principal Component Analysis
% Scores of the US cities based on 9 different categories.
% Which factors of city life make a difference?
load cities
%%
% Get a first quick impression about the data
boxplot(ratings,'orientation','horizontal','labels',categories)
% You could also use 'plot' to compare pairs of variables,but there would be 36 two-
variable plots!
%%
% Get a quick idea of correlations
corr(ratings)
%%
% Standard deviations for at least 2 categories are substantially different. We need to
standardize the data. Divide the data by the corresponding standard deviations
stdr = std(ratings);
sr = ratings./repmat(stdr,329,1);
% The standardized rankings are now in variable 'sr'
boxplot(sr,'orientation','horizontal','labels',categories)
%%
% Now find the principal components!
[coefs,scores,variances,t2] = princomp(sr);
%%
% First, look at first 3 vectors of principal component coefficients
c3 = coefs(:,1:3)
%%
% Component scores (variable 'scores') are the original data
% mapped into the new variables
% Projection on the first two (most significant) principal components:
plot(scores(:,1),scores(:,2),'+')
xlabel('1st Principal Component');
ylabel('2nd Principal Component');
% Note the outlying points
%%
% This command allows you to click on data points and shows their labels
gname(names)
%%
% Remove largest cities from the data
metro = [43 65 179 213 234 270 314];
names(metro,:)

%To remove these rows from the ratings matrix
rsubset = ratings;
nsubset = names;
nsubset(metro,:) = [];
rsubset(metro,:) = [];
size(rsubset)
%%
% Using the 'variances' output, calculate the percent of variance in the data explained by
each principal component
percent_explained = 100*variances/sum(variances)
```

```
pareto(percent_explained)
xlabel('Principal Component')
ylabel('Variance Explained (%)')
%%
% Using the 't2' output, find the most extreme observation
[st2, index] = sort(t2,'descend'); % Sort in descending order.
extreme = index(1)
names(extreme,:)
%%
% Visualize the results of the principal component analysis
biplot(coefs(:,1:2), 'scores',scores(:,1:2),...
'varlabels',categories);
% Each of the nine variables is represented in this plot by a vector, and
% the direction and length of the vector indicates how each variable
% contributes to the two principal components in the plot.
axis([-.26 1 -.51 .51]);
```